Estimating a nonparametric homothetic S-shaped production relation for the US West Coast groundfish production 2004-2007,

National Marine Fisheries Service Productivity Workshop.
Southwest Fisheries Science Center.
Santa Cruz, CA June 11-12, 2012

O. B. Olesen, J. Ruggiero and and A. T. Mamula

06.11.2012

Contributions:

- The VRS - DEA model is developed with a maintained hypothesis of convexity in input-output space.
- This hypothesis is not consistent with production theory that posits an S-shape for the production frontier.
- Consequently, measures of technical efficiency assuming convexity are biased downward.
- We provide a more general DEA model that allows for an explicit estimation of the S-shape (the scaling law).
- Simplification: One output multiple inputs and an input homothetic technology.
- A homothetic production function is a monotonic transformation of a linear homogenous production function (introduced in Shephard (1953))
- In geometric terms, any input isoquant from a homothetic production function is simply a radial rescaling of the unit isoquant.
- Input homotheticity allows for an aggregation of inputs

Contributions:

- The VRS - DEA model is developed with a maintained hypothesis of convexity in input-output space.
- This hypothesis is not consistent with production theory that posits an S-shape for the production frontier.
- Consequently, measures of technical efficiency assuming convexity are biased downward.
- We provide a more general DEA model that allows for an explicit estimation of the S-shape (the scaling law).
- Simplification: One output multiple inputs and an input homothetic technology.
- A homothetic production function is a monotonic transformation of a linear homogenous production function (introduced in Shephard (1953))
- In geometric terms, any input isoquant from a homothetic production function is simply a radial rescaling of the unit isoquant.
- Input homotheticity allows for an aggregation of inputs

Production Technology (1)

A vector of s inputs $X = (x_1, \ldots, x_s)$ is used in the production of one output Y

The input set: $L(Y) = \{X \in \mathbb{R}^s_+ : X \text{ can produce } Y\}$

Isoquants: $IsoqL(Y) = \{X : X \in L(Y), \lambda X \notin L(Y), \lambda \in [0,1)\}$.

Only one output implies that we can define a production function as

$$\phi(X) = \max\{Y : X \in L(Y)\}$$

The input distance function is then defined as

$$D_I(Y, X) = \max\{\gamma : X/\gamma \in L(Y)\}$$

Farrel technical efficiency

$$F_I(Y, X) = \min\{\gamma : \gamma X \in L(Y)\}$$

where $F_I(y, x) = D_I(y, x)^{-1}$

Production Technology (2)

We assume that production is homothetic.

Definition

A production function $\phi(X)$ is homothetic

$$Y = \phi(X) = F(g(X))$$

where $F() : R_+ \rightarrow R_+$ is monotonic and $g(\lambda X) = \lambda g(X)$ i.e. $g()$ is positive homogeneous of degree one and continuously differentiable. $g()$ is denoted the kernel function.

A homothetic production function can be represented as a production process whereby the input vector X can be aggregated into a one dimensional input index $g(X)$, i.e. output is determined from the level of aggregate input.

Production Technology (3)

Proposition

Assume a homothetic technology with one output. The input distance function evaluated at $(1, X)$ is equal to aggregate input defined from the core function in the homothetic production function multiplied by a constant, i.e.

$$D_I(1, X) = k \times g(X), k \in \mathbb{R}_+$$

- The dimensionality of DEA models can be reduced under the assumption of homotheticity using $g(X)$
- We span the production technology from the input set associated with the unit isoquant $L(1)$

$$L(Y) = H(Y)L(1),$$

- $H(Y)$ is the possibly S-shaped (inverse) scaling function.
- We choose the unit isoquant for convenience only.

A conditional estimator of the Base Isoquant

- A simple "conditional" estimator used in the simulation of the base isoquant is estimated using

$$\widehat{\theta}_C^{LC}\left(X_i, Y_{base}\right) = \begin{cases} \min & \theta - \varepsilon\left(1, \ldots, 1\right) s \\ s.t. & \theta X_i - \sum_{j=1}^{n} \lambda_j X_j - s \quad = \quad 0 \\ & \sum_{j=1}^{n} \lambda_j \quad = \quad 1 \\ & \lambda_j = 0 \text{ if } Y_j < Y_{base} \\ & \lambda \in \mathbb{R}_+^n, s \in \mathbb{R}_+^s \end{cases}$$

- $i = 1, \ldots, n$, where $Y_{base} = 1$
- The estimator $\widehat{L}_C^{LC}(1)$ of the input set is derived as

$$\widehat{L}_C^{LC}(1) = Conv\left[\widehat{\theta}_C^{LC}\left(X_1, 1\right) \times X_1, \ldots, \widehat{\theta}_C^{LC}\left(X_n, 1\right) \times X_n\right] + \mathbb{R}_+^s.$$

Choosing the "best" isoquant

As base isoquant we now choose the specific output level which performs well according to:

1. The distribution of the angle coordinates of the data points spanning the estimator of the isoquant should mimic the uniform distribution (on the empirical support of the angles).
    - The area between the empirical distribution and a uniform distribution is used as a measure of the deviation.
2. A large number of observed data points should be located on the estimator of the isoquant.
3. As many as possible of the data points should be projected to the envelopment, i.e. should be located within the cone spanned by the points that span the isoquant.

Implementing the selection process (1)

- Estimate all $n \times n$ conditional scores providing $\widehat{\theta}_C^{LC}(X_i, Y_j), j, i = 1, \ldots, n$, (missing, if $Y_i < Y_j$).
- For a given observed output level Y_{j_o} keep the input vectors X_i if $\widehat{\theta}_C^{LC}(X_i, Y_{j_o}) = 1$.
- Collect the angles in input space corresponding to observations on the estimated isoquant at output level Y_{j_o}.
- With two inputs (the simulations) there is only one angle in the polar representation of the input vectors.
- Sort the angles and plot the empirical distribution against the uniform distribution
- Calculate the deviation of this empirical distribution from the uniform distribution

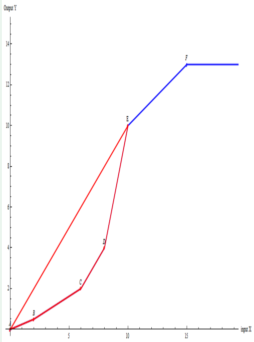
A piecewise linear S-shaped estimator (1)

- Assume that the boundary of T^S is S-shaped.
- Divide the input axis into two parts $[0, x^*]$ and $[x^*, \infty)$ where the production function is convex and concave.
- Marginal product is monotonically non-decreasing (non-increasing) in $[0, x^*]$ (in $[x^*, \infty)$).
- The convex hull estimator $\widehat{T}^{BCC}$ works well above the inflection point x^*.
- We remove or "dig out" the part of the estimator $\widehat{T}^{BCC}$, that violates the S-shape.
- We dig out a certain convex hull of observed data point that satisfies the following:
  - the convex hull is spanned by points below (and on) the inflection point.
  - the convex hull is constructed such that no point is located above the frontier (i.e. in the interior of the hull).

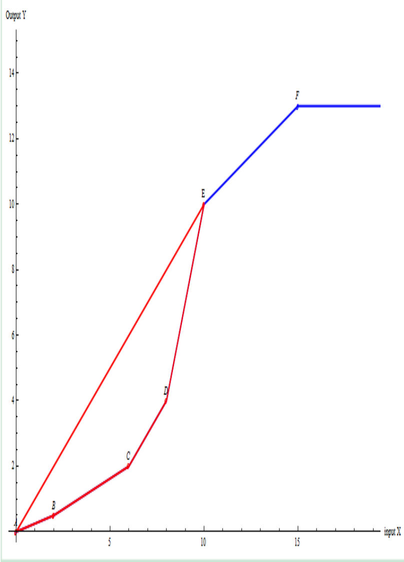
A piecewise linear S-shaped estimator (2), a simple illustration.

The Figure illustrates, where A, E and F are BCC-efficient and E is most productive scale size (mpss).

We are looking for an estimator $\widehat{T}^S \equiv \widehat{T}^{BCC} \setminus \widehat{T}^{Dig}$ of the PPS with an S-shape, $\widehat{T}^{BCC} = Conv\,(A, E, F) + \mathbb{R}_+ \times \mathbb{R}_-$, and $\widehat{T}^{Dig} = Conv\,(A, B, C, D, E)$



A piecewise linear S-shaped estimator (2), a simple illustration.



Non uniqueness of the piecewise linear S-shaped estimator.

- The determination of this inverted convex hull is unfortunately not unique.
- Notice, L being below the extension of the facet CD.
- Cannot include both L and A, B, C, D and E on the frontier.
- Can include L as on the frontier if we remove C from the frontier (the dashed convex hull).
- Either C or L is efficient, but not both.

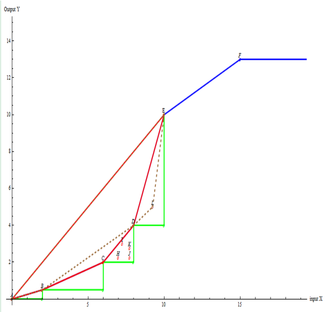
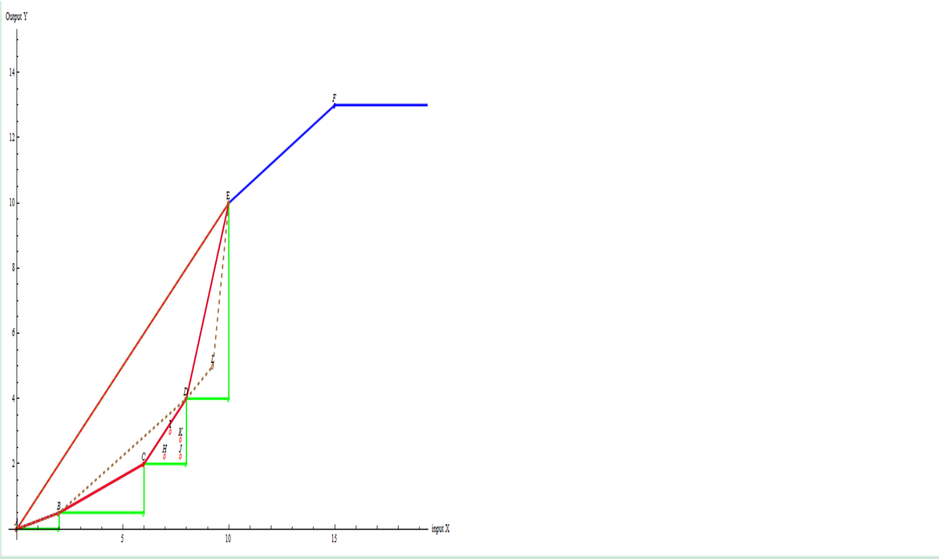


Illustration of the BCC- and the S-shaped frontier

Non uniqueness of the piecewise linear S-shaped estimator.



A general estimator of the S-shape (1)

- Maximize the number of FDH efficient points on the S-shaped frontier.
- At the same time estimate the inflection point.
- Testing a given point as a candidate for the inflection point involves several conditions:
  - the marginal products along the facets from the origin to the inflection point must be increasing,
  - the marginal products must be non-increasing on facets above the inflection point and
  - all the points have to be on or below the frontier.

For FDH-efficient points below a candidate (X_n, Y_n) for the inflection point

$$\begin{array}{rcl} (X_j, Y_j) & \in & R_+^2, j = 2, \ldots, n+1, \\ (X_1, Y_1) & = & (0,0) \text{ and} \\ \dfrac{Y_{j+1}}{X_{j+1}} & > & \dfrac{Y_j}{X_j}, j = 1, \ldots, n-1, \dfrac{Y_{n+1}}{X_{n+1}} < \dfrac{Y_n}{X_n} \end{array}$$

A general estimator of the S-shape (2)

- We are looking for a convex shape as a graph through a subset of the points $(X_j, Y_j), j = 1, 2, \ldots, n+1$, starting at $(X_1, Y_1) = (0, 0)$ and ending at the estimator of the inflection point (X_n, Y_n).
- This problem resembles the so-called traveling salesman problem (TSP)
- TSP looks for the shortest path through a set of points, never visiting a point more than once and returning to the starting point.
- In our problem the length of the path does not matter and it is not required that we return to the starting point.
- But it is required that we start at point 1 and end at point n.
- Secondly, we do not require that the path covers all points. In fact, we expect only a subset of points to be covered.
- But we maximizes the number of points visited upto and including the inflection point.
- Thirdly, the $(i+1)\prime th$ edge is required to have a larger marginal product compared to the $i\prime th$.
- Finally, all points have to be on or below the frontier

Simulation: recovering the true S-shaped technology (1)

- Aggregated input indices: $\left(\theta_c^{LC}(X_l, Y_{750})\right)^{-1}, \forall l$
- Remove the observations with positive slacks present (181 observations) and observations that are FDH inefficient (732).
- Data on 88 observations of which 16 (71) are above (below) mpss.
- We estimate the BCC efficiency scores based on the sample of observations from 72 to 88 (the concave part)
- We endogenously determine which point below mpss is the inflection point.

Simulation: recovering the true S-shaped technology (2)

Inflection Point	71	70	69	68	67	66	65	64	63
The concave part up to mpss	2	3	2	3	4	5	6	7	6
The convex part	13*	–	17	21	25	27	17	–	25
The frontier up til mpss	–	–	19	24	29	32	23	–	31
Notes: – indicate that the binary LP is integer infeasible									
* indicate that the cutting procedure has stopped after 50 cuts									

Table 1: The number of points one the different parts of the frontier with different choices of the inflection point.

Simulation: recovering the true S-shaped technology (3)

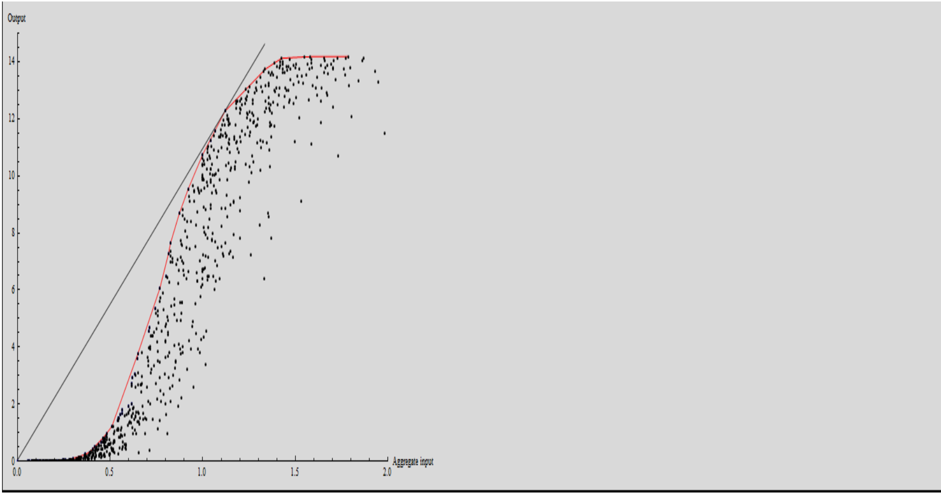
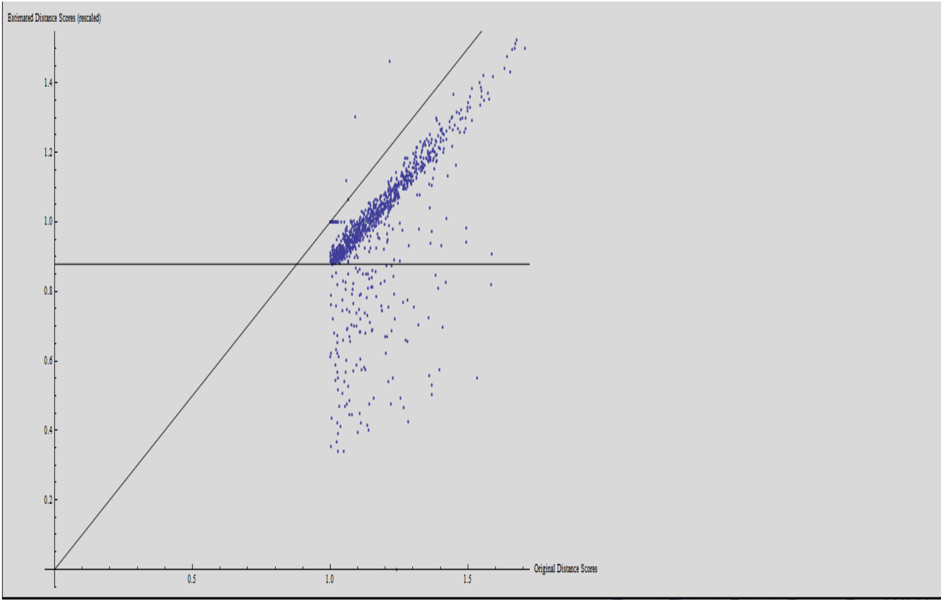


Figure 4. The estimated S-shape and FDH efficient points below and above mpss.

Simulation: Estimated and True Distance Scores



Application to US West Coast Groundfish Production (1).

- We now turn to an application of the proposed estimation of a possibly S-shaped production relationship the West Coast Groundfish production 2004-2007.
- One main focus for our analysis is:
  - Can we find indications of a purely increasing return to scale segment of the production of harvest of fish?
- Our estimation procedure relies on a maintained hypothesis of an output homothetic production.
- We avoid any specific function form of the production relation

Application to US West Coast Groundfish Production (2).

- The application of the model is based on a data set with catch records from fishery in the pacific ocean during the period 2004-2007.
- The entire catch record for each trip from each vessels is used.
- Primarily, based on the condition of the fishery the catch is aggregated into the following 2 outputs:

1. i) the catch in lbs. of dover sole, sablefish and thoryhead rockfish (generally encountered at depths greater than 200 fathoms), ii) the catch in lbs. of various rockfish species which are harvested mainly along the continental shelf at depths between 75 and 150 fathoms
2. iii) the catch in lbs. of a nearshore mix species aggregate (mainly petrale sole, rex sole, sanddabs and other flatfish), iv) the catch in lbs. of California halibut which is harvested in shallow water less than 75 fathoms.

Application to US West Coast Groundfish Production (3).

- We have chosen to define a DMU as a vessel in a specific year.
- Hence, we sum the catch in these two categories and the variable input "days at sea" for fixed vessel for each of the four years.
- The total data set comprises 192 vessel-year combinations.
- We have removed five outlier observations which brings us to a final sample size of 187.

Choice of output isoquant (1)

- Using the guidelines for how to choose the input level with the most useful information we will now use the conditional estimator relative to a given input level x and only include output vectors from observations with an input level below or equal to this x.
- We look for a specific isoquants (a x level) where
- i) we have many observed points on or just below the isoquant,
- ii) where the points are spread out evenly along the full isoquants, and
- iii) where as many of all the other data points are radially projected to the envelopment of these points, i.e. are located in output space within the cone spanned by the points that spans the isoquant.

Choice of output isoquant (2)

- We sort the data on the variable input "days at sea" and estimate 187 output oriented scores for each isoquant for each of the 187 input levels;
- We thus obtain 187 output oriented scores for each isoquant.
- If the output oriented score has additional slack in any of the output dimensions the score is assigned the value "missing".
- We identify for each potential base isoquant those points that span the conditional isoquants
- For this application, we consider the three criterias mentioned previousy for choosing our base isoquant.

Choice of output isoquant (3)

1. We are looking for the particular isoquant with as many observation on the frontier as possible.
2. We search for sets of points with an empirical distribution of the angles as close as possible to a uniform distribution.
3. We want the spanning points to span as large a cone as possible in output space compared to the cone spanned by the full sample of data points.

- Figure 4 illustrates the first criterion.
- On the horizontal axis we have the 187 different estimators of the output isoquants.
- On the vertical axis we measure the number of data points on or close to the estimated frontier.
- The red curve includes the number of points with an output oriented score in the band [1,1.001].
- Isoquant 60 seems to be performing well on this criterion.

Choice of output isoquant, criterion 1

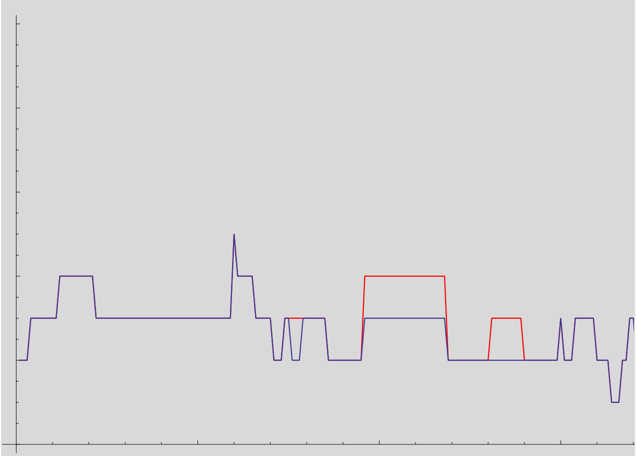


Figure 4: The 1. criterion: The number of data points located on (blue) or on or close below (red) the estimated frontier.

Choice of output isoquant, criterion 2-3

- Figure 5 illustrates the second criterion.
- On the vertical axis we measure the deviation of the output mix from the uniform distribution.
- Finally, Figure 6 illustrates the third criterion
- We have counted the number of observation being in output space inside the cone spanned by the output vectors spanning each of the 187 estimators of the isoquant.
- Again we see that isoquant 60 is a promising candidate, but the isoquants 102-115 do seem promising too

Choice of output isoquant, criterion 2-3

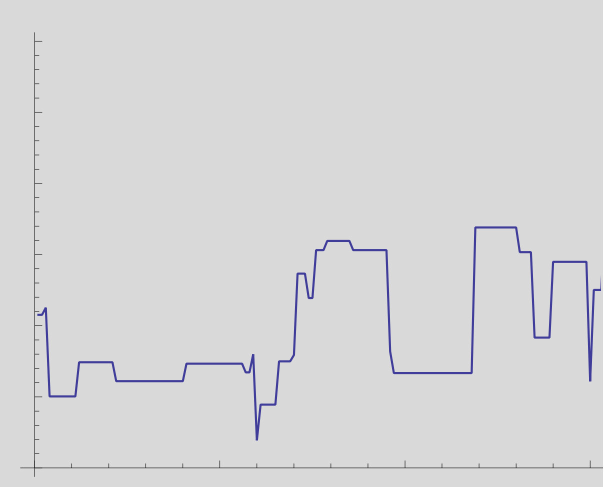


Figure 5: The 2. criterion. The deviation of output mix from the uniform distribution.

Choice of output isoquant, criterion 2-3

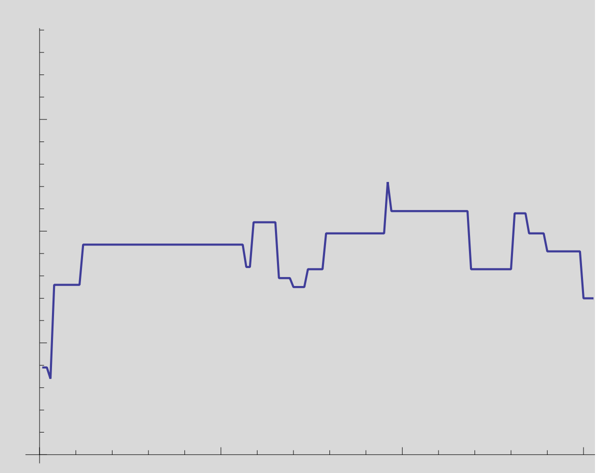


Figure 6. The 3. criterion: The number of sample points projected to the estimator of the isoquant (without slacks).

The estimated piecewise linear S-shape.

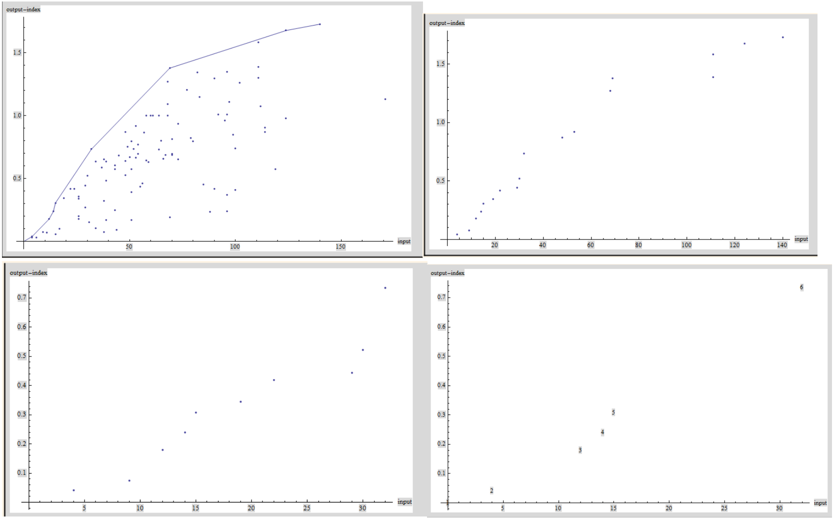


Figure 7a,b,c,d. Graphical illustrations of the construction of the S-shaped input consumption function

The estimated piecewise linear S-shape (1).

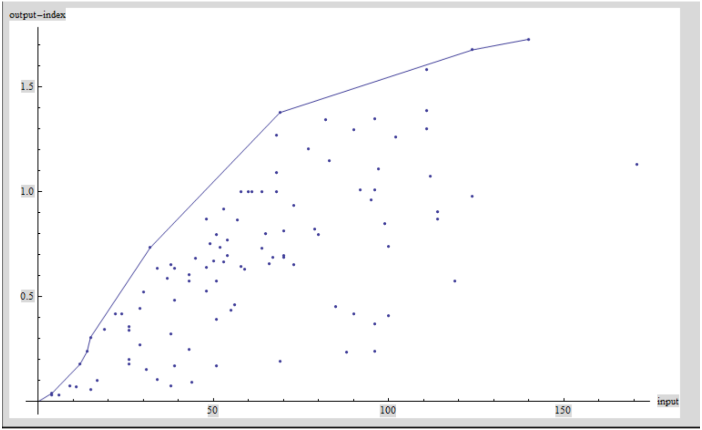


Figure 7a. The estimated S-shape

The estimated piecewise linear S-shape (2).

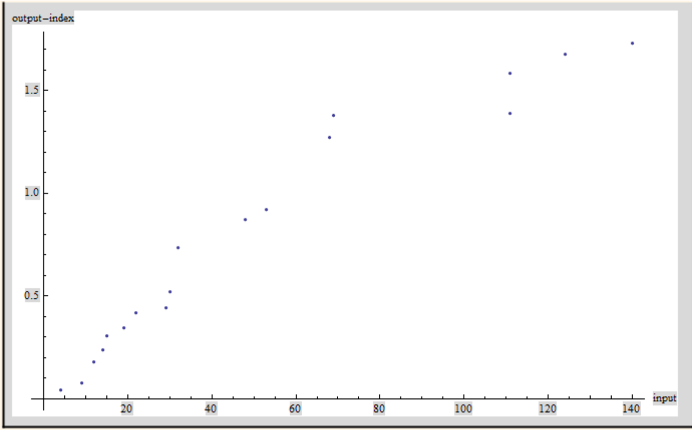


Figure 7b. The FDH efficient observations, (input in $[0, 140]$)

The estimated piecewise linear S-shape (3).

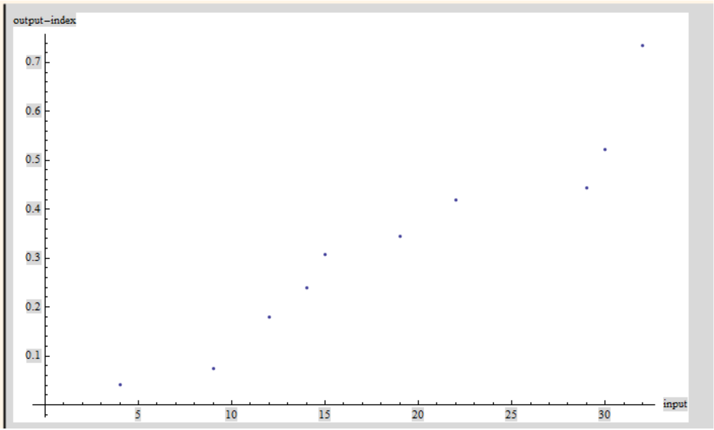


Figure 7c. The estimated S-shape (input in $[0, 30]$)

The estimated piecewise linear S-shape (4).

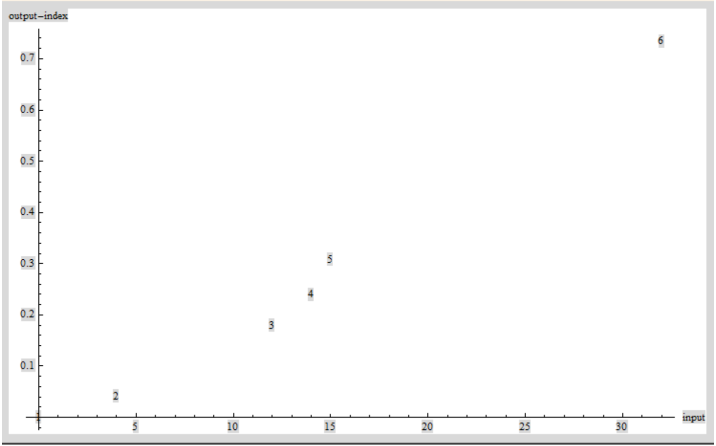


Figure 7d. The convex-concave part of the S-shape (input in $[0, 30]$)

Conclusion and Summary

- In this paper we have outlined an approach that allows for an estimation of efficiency from an S-shaped technology for the multiple inputs and one output case.
- To simplify, we have assumed that the technology is input or output homothetic.
- This assumption has allowed us to split the estimation procedure into two parts,
  - i) an aggregation procedure based on the structure of input or output homotheticity, and
  - ii) a joint estimation of the inflection point and a piecewise linear S-shaped structure for one aggregated input and one output/ aggregated output and one input.
- We are currently working on:
  - extending this procedure to other non-homothetic structures
  - extending this approach to a multiple input and multiple output situation